SAN FRANCISCO PUBLIC LIBRARY
3 1223 11169 5251

bart impact

working paper

TRANSIT PASSENGER COUNTS:
ERRORS, CONTROLS, ANALYSIS

By Andrew Ungar

METROPOLITAN TRANSPORTATION COMMISSIO
HOTEL CLAREMONT
BERKELEY, CALIFORNIA 94705

D
REF
388.42
B2813tp

TRANSIT PASSENGER COUNTS: ERRORS, CONTROLS, ANALYSIS

By Andrew Ungar

METROPOLITAN TRANSPORTATION COMMISSION
HOTEL CLAREMONT
BERKELEY, CALIFORNIA 94705

Document No. WP 11-1-75
Prepared by the Metropolitan Transportation Commission
for the U. S. Department of Transportation
and the U. S. Department of Housing and Urban Development
Under Contract DOT-OS-30176, Task Order No. 1
NOVEMBER, 1974

3 1223 11169 5251

CONTENTS

		PAGE
	SUMMARY	1
I.	INTRODUCTION	3
II.	METHODS OF CONDUCTING TRANSIT PASSENGER COUNTS	5
	A. Counting Methods	5
	B. Prior Work in Seattle	7
	C. Field Procedures	12
	D. Comments on Field Methodology	15
III.	RESULTS AND ANALYSIS	17
	A. Analytical Approach	17
	B. Results	19
IV.	IMPACT ANALYSIS AND FUTURE DATA COLLECTION	29
	A. Sample Sizes	29
	B. Quality Control	30
	APPENDIX A - A model for Detection and Estimation of Changes in Transit Patronage	33
	APPENDIX B - Methods for Computing Type II Errors	42
	APPENDIX C - Written Instructions to Transit Count Teams	45

Digitized by the Internet Archive
in 2015

https://archive.org/details/transitpassenger1974unga

SUMMARY

This paper reports work performed by the staff of the Metropolitan Transportation Commission to determine the accuracy of two methods of counting bus riders (by placing observers on board the buses, and by stationing them beside the roadway), to devise procedures that will produce the most accurate data at a reasonable cost, and to provide estimates of the error inherent in the adopted procedures.

Bus riders were counted by both methods at selected locations on a screen line in San Francisco, as part of a larger program of vehicle and traveler counts performed by MTC's BART Impact Program with financial and technical assistance by the California Department of Transportation.

The use of two counting methods permitted estimation of counting errors and of data variability. Based on those estimates, it was concluded that the standard counting procedure, using roadside observers, contributed relatively little to the overall variability of the data. The largest source of variation was the normal fluctuation of ridership between buses that every transit rider is familiar with.

With the observed magnitudes of errors, five-day samples taken in two observation periods would be required to permit the detection, on the average, of a 10 per cent change in total daily ridership on a single street or highway crossing the

screen line. It is expected that smaller differences in total ridership in an entire corridor can be detected by judicious pooling of data from several counting stations.

The check-counting procedure also pointed up problems of quality control in the training and supervision of field crews, and in data management procedures. A program of quality control, growing out of the experimental procedure and the associated analysis, has been designed and has been implemented in subsequent counts of bus passengers.

I. INTRODUCTION

Many of BART's impacts may prove to be rather small, and may therefore be difficult to detect and measure against a background of random variation in the variables used to measure the impacts, combined with on-going changes in other variables. All parts of the BART Impact Program need highly accurate measurement methods and well-founded estimates of measurement error, in order to assure that relatively small impacts can be detected, and that levels of statistical significance can be attached to statements about impact.

This problem has been addressed in the collection of data on highway traffic volumes, including counts of both vehicles and people. Concern for the accuracy of counts of bus riders led to the work whose results are described herein. The objectives of this work were to assess the accuracy of two possible methods of counting bus riders, to devise procedures that represent an optimum combination of accuracy and and economy, and to provide estimates of the error inherent in the adopted procedures.

There was an additional interest, of equal importance, in using the accuracy assessment procedures to help identify problems of data quality control. The assurance of data quality by means of controls on field procedures and data management is essential, and can best be addressed simultaneously with the determination of how many observations are required to detect changes in transit patronage, say during peak periods at various screen lines.

All other factors being equal, the ability to detect a change of a given magnitude at a given level of significance, depends on the variance of the observations -- the smaller the variance, the smaller the required sample size. The variance of the observations may be regarded as having two components, the human counting error, and the day-to-day variation of patronage. The observer can do nothing about the variation of patronage, but there are many things that can be done (or neglected) to affect counting error. At a price (supervision, quality checking procedures, etc.), counting error can be reduced to a small fraction of the total variance. The offsetting benefit is, of course, that less data need be collected.

II. METHODS OF CONDUCTING TRANSIT PASSENGER COUNTS

A. Counting Methods

The BART Data Acquisition System (DAS) will, when it is operational, be capable of providing tallies of all trips taken, by station of origin, station destination, and time of day. The information will be extracted from the traveler's ticket by the fare computation equipment at the exit gate, and will come as close to absolute accuracy as computer technology can make it.

There are no comparable means of counting bus and streetcar passengers. Any method that is used to count passengers crossing a screen line must employ people to do the counting, and, depending on the method and the people, is subject to some greater or lesser degree of error.

In counting bus riders crossing the San Francisco screen line, two methods were used. Observers were stationed at the roadside and on board the bus.

Roadside counters stand at a bus stop immediately "upstream" or "downstream" from the screen line. All the buses come to them, so that, except for unusual circumstances, they can count the passengers on every bus. However, buses stop for a relatively short time; when more than one bus is at the stop there is not always a good vantage point from which to make

a count; at a busy stop, it is difficult to count arriving passengers at a downstream stop before the bus unloads, or departing passengers at an upstream stop after the bus has loaded. All of these factors contribute to counting errors.

Onboard observers ride buses back and forth across the screen line, counting passengers as they go. They have a longer time to make their count than the roadside observers do and, except in very crowded buses, can move down the aisle to observe all parts of the bus. For these reasons, one would expect onboard counting to be more accurate than the roadside method. However, experience shows that a trio of onboard observers may be able to count only one-third as many buses as a roadside crew of the same size, depending on the average headway at the screen line. So onboard counting, though more accurate, is more expensive than the roadside method.

Both methods are vulnerable to recording errors, misunderstanding of instructions, and the hazard of poor motivation leading to slipshod counts, all of which need to be met by active and meticulous supervision.

Two other possible counting methods employ the services of transit system personnel. One is the use of bus drivers to make counts of their own passengers; the other is the use of transit system employees who spend much of their time

making counts for use in planning and operations.

The apparent advantage of bus drivers to make counts is the potential for greater accuracy of onboard counts without the expense of hiring special crews to do the counting. Such counts, however, are difficult for drivers to make on crowded buses with tight schedules. Also, there is always the complication of administrative problems and extra costs in requiring drivers to add chores outside their normal responsibilities.

The transit companies' special survey crews, while they use the roadside counting method, are assumed to be more skilled and responsible than the temporary crews hired for a single survey. The special crews comprise only two or three people, which is not enough to do an extensive survey over a short time period. However, their experience can be utilized for spot checks on the quality of the temporary crews' work, which has been done in a survey subsequent to the one reported here.

B. Prior Work in Seattle

An approach to the use of a control counting method to assess the accuracy of other counting methods was reported from Seattle* prior to our work in San Francisco. The report described a short experiment, in connection with an Urban Mass Transportation Demonstration Project, in which counts by bus drivers on one day and by roadside observers on another were compared to a control method, namely counts by onboard

*"Evaluation of Procedures for Collecting Data on Blue Streak and Local Routes in Seattle, Washington", unpublished paper prepared on an UMTA grant by Richard M. Michaels and Ross Adams, then (1972) both of Northwestern University, Evanston, Illinois.

observers. The principal conclusions of interest to us in the Seattle paper were that:

(1) Bus drivers, on the average, overcounted by a significant amount as compared to onboard observers, and their overcounting increased as the number of passengers increased.

(2) There was no significant difference, on the average, between the counts of roadside and onboard observers.

(3) The variance of the roadside counting method was greater than that of the onboard control.

(4) The error associated with the roadside counting method was too great to permit the method to be used for detecting small differences in ridership.

As was previously noted, bus drivers were not used in San Francisco to make any counts, so the corresponding finding, that bus drivers overcounted, was not checked.

To anticipate results a little, conclusions (2) and (3) were confirmed in San Francisco. However, re-analysis of some of the Seattle data, and observations in San Francisco, showed that "no significant difference" between the two counting methods cannot be automatically assumed. Depending on circumstances, which must be attributed to human factors, biases

(systematic over- or undercounting) may creep into the results.

The final conclusion, that the error associated with the roadside counting method is too great to detect small differences, resulted from some omissions and oversights in the statistical analysis:

1. Although in some instances onboard checkers rode buses in pairs, thereby providing an opportunity to make an independent estimate of the error of the onboard counting procedure, no such estimate was reported.

2. The conclusion about the overlarge error of roadside counts was based on the distribution of differences between counts made of the same bus by the two different methods, roadside and onboard counting. The variance of a difference is the sum of the variances of the numbers being differenced.* Thus, what was reported as the error of the roadside method was actually the sum of the errors of the two methods combined. If the independent estimate of the variance of the onboard counting procedure had been computed, it could have been subtracted from the variance of the differences to obtain an estimate of the variance of the roadside counting procedure alone.

*Provided the covariance is zero. We are justified in assuming that to be the case here, that is, that the two observers counted independently of each other.

3. The statement about the error being too large ignores the general principle that, given a variance of a certain magnitude, a difference of any size can be detected provided a large enough sample is taken. There are, of course, always practical limitations on the size of sample that can be taken (cost, time, span of control), but one can always estimate how large an adequate sample might be.

4. Again, to anticipate results somewhat, it turns out that much the largest contributor of variability in the data, and hence the determinant of detectable differences, is the variability of ridership from bus-to-bus. In this context, the error of the counting method is a relatively small part. This is a factor that was not considered at all in the analysis of the Seattle data. It may not be intuitively clear why the between-bus variability should matter, or indeed why one is permitted to use that information at all (why not use instead the peak-period total and ignore between-bus variation?). The reason, in brief, is that if one can show justification for regarding the individual bus counts as coming from a normal distribution, the additional information provided by the known properties of the distribution can be used to reduce the required

sample size as compared to that required when the measure of change is the total count in the peak period. The analysis of the distribution of between-bus variation and its effect on sample sizes may be of interest mainly to statisticians; a detailed discussion of the analytical results has therefore been relegated to Appendix A.

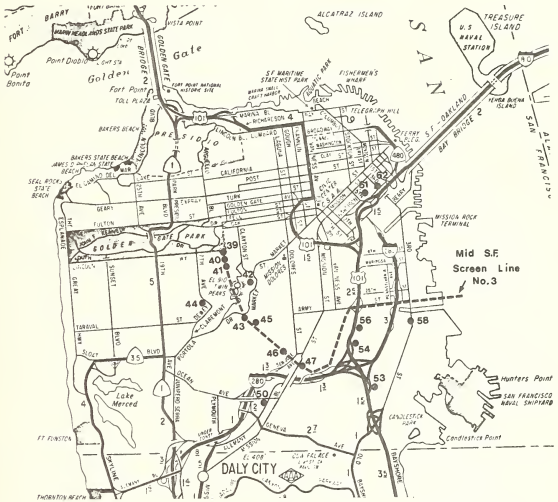
C. Field Procedures

The transit occupancy count program was designed to obtain independent passenger counts of the same bus at the same point by two kinds of observers. The transit occupancy counts were part of a larger program of counting that also included vehicular traffic volumes, vehicle type classification counts and passenger car occupancy counts. Besides the Muni (San Francisco Municipal Railway), transit occupancy counts were made on the Mission Street jitneys and the commuter buses from points south of San Francisco. The mid-San Francisco screen line, across which counts were made, is shown in Figure 1. The numbered points are stations at which both roadside and onboard counts were made, all of Muni patronage. Those stations shown at some distance from the screen line are last (or first) stops before (after) the screen line is crossed, usually associated with express lines. The vehicles observed at Station 44, in the Twin Peaks tunnel, were streetcars. All others were diesel buses or trolley buses.

Counting crews worked in two shifts per day, an AM and a PM shift. On each shift, there were two roadside crews per station, one on the inbound side, and one outbound. There were three persons per crew, one making passenger counts and calling them out to a second, who recorded them as well as the time and vehicle identification data. The third person, in rotation, was on break. Passenger counts were made

Figure 1.

VEHICLE OCCUPANCY COUNTING STATIONS ALONG MID-SAN FRANCISCO SCREEN LINE



Base information from AAA

of all buses arriving at or departing from a station, depending on whether the bus had just passed or was about to cross the screen line. Onboard crews also consisted of three persons, one crew per shift. With a few exceptions to be noted, they worked individually, boarding buses singly. A crew member was supposed to board a bus at the last stop before the screen line, make his count after departure, and alight at the first stop. He then boarded the first bus going in the opposite direction for which there was not already an onboard counter -- and so on throughout the shift. In addition to the counts and vehicle identification data, the time of crossing the screen line was recorded, to help in matching with roadside data. Depending on frequency of service at a station, the shuttling process was able to sample from one-third to almost all of the buses.

Provision was made for an independent check of onboard counting precision. At a different station on each of three days, onboard counters were assigned to ride vehicles in pairs, but to make independent counts and to record their results separately.

Crews were trained in the prescribed counting and data-recording methodology through instruction and several days of field work. They were also provided with written instructions. For each pair of crews at a station, one was appointed team captain. The captain was responsible for supervision at the station. In addition, CALTRANS* personnel acted as roving

*Department of Transportation of the State of California

field supervisors. MTC's BART Impact Program staff maintained daily contact with the onboard crews, receiving their comments and observations, and adjusting procedures accordingly. Examples of written instructions to the field crews are presented in Appendix C.

Matched roadside/onboard counts were made at 15 stations, one day at each, and two days at a sixteenth station. A full day of counts was divided into six sets, of which three were inbound (toward the CBD) and three outbound. In each direction, the three sets were AM peak (6 to 9 AM), offpeak (9 AM to 3:30 PM) and PM peak (3:30 to about 6:30 PM). Some sets were incomplete, for a variety of scheduling reasons. Of a possible 102 data sets (17 station-days x 2 directions x 3 periods), usable data were obtained for 84. Counts were made on seven different days.

D. Comments on Field Methodology

There were many instances in the data where discrepancies between onboard and roadside records were such as to require some other explanation than counting error. Examples are where one crew reports an empty bus, whereas the other reports riders; or standees on a full bus versus no standees; or a nearly empty bus versus a crowded bus.

There are many possible explanations for the discrepancies. Motivation and the consequent degree of conscientiousness are always a question. While this is not ruled out as a

factor, the field supervisors reported a generally satisfactory level of performance.

The likeliest explanation for most of the discrepancies is that the paired counters were often counting occupants at different points on the run, that is, before and after stops at which the ridership changed substantially. These errors may have arisen because of misunderstood instructions, or ambiguous or incomplete instructions.

An additional source of error may be the occasional inability to observe properly. In the case of the roadside crews, at crowded, busy stations, if two or more buses arrived close together, there may not have been time to do an accurate count or a good vantage point from which to do it. In the case of an onboard counter boarding a crowded vehicle, he may not always have found it possible to move to a position where he could easily see all the standees.

There were also obvious recording errors; counts incorrectly entered, erroneous vehicle and route numbers, disagreements over count times (sometimes due to unsynchronized watches).

The immediate effect of the observed data problems was the need to expend significant effort, after the fact, in examining and correcting the data. In some instances it was obvious, for example, that one of the observers had inverted the order of the digits of a bus's identification number. Such data could be corrected and were used. In other cases,

the data could not be salvaged.

The longer-range effect of the problems was to cause a re-evaluation of field procedures, and the design of controls to assure better data, to be implemented in subsequent surveys. There is a residue of error that can never be completely eliminated, deriving from human psychological factors (how people see, fatigue, boredom, etc.) and the physical conditions in which counts are made. However, errors deriving from improperly implemented procedures, and some recording errors, can be detected and corrected through appropriate controls. The control procedures resulting from the re-evaluation are described in Section IV.

III. RESULTS AND ANALYSIS

A. Analytical Approach

The analytical approach was as follows:

1. In the analysis comparing the two counting methods, the variable of interest was the difference (roadside - onboard) of the two counts of riders on the same bus. These data were used to test the hypothesis that, on the average, there was no significant difference between the counting methods. As noted previously, there were 84 sets of data available to test. They were also used to estimate the error of the two methods combined.

There were also available two sets of data in which pairs of onboard observers counted the same bus, independently of each other. These data were used to compute an estimate of the error associated with the onboard method alone. That error was then subtracted from the error of the two methods combined, to obtain an estimate of error for the roadside method alone.

In all cases, to test the hypothesis of no difference between counting methods, the 2-tail t-test was used, with a significance level of 0.1.

2. The analysis aimed at estimating sample sizes required to detect changes in patronage of various magnitudes used the roadside counts from the station at which counts were made on two days. The standard approach of using the operating characteristic curve* of the t-distribution was followed, although in this application some special problems were encountered which are discussed in Appendices A and B.

3. The problems with the quality of some of the data have already been discussed. In the analyses, a relatively small number of extreme values were excluded. Observations were judged case-by-case, and a difference was considered to be extreme only when

*Described in many statistics texts. See, for example, Albert H. Bowker and Gerald J. Lieberman, Engineering Statistics, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1959.

there was no plausible explanation for it except gross error. Some differences as large as 15 were retained nevertheless, when there was no obvious reason to discard them. A typical example of excluded data is one case where the roadside count was 18 and the onboard was 0. In another case, both observers recorded all seats taken, but the roadside observer saw no standees, whereas the onboard standee count was 21.

Two viewpoints are possible with respect to the excluded data. One is that whatever errors occurred, gross or otherwise, they were inherent in the counting methodology, and should therefore be retained. The other is that most gross errors are avoidable, that steps can and should be taken in future surveys to avoid them, and that a more realistic estimate of the variabilities of the contrasted counting methods is obtained by excluding the extreme data. The second viewpoint was adopted.

B. Results

Summary statistics associated with the t-tests of the 84 sets of roadside/onboard differences are presented in Table 1. Summary statistics for the duplicate onboard counts are in Table 2.

1. Comparison of Two Methods of Counting

The 84 t-tests confirmed that, on the average, there is no significant difference between the counting methods. The evidence for this conclusion is mixed with indications of occasional quality control problems. It consists of the facts that of the 84 tests, 64 showed no significant difference, whereas 20 were significant; and of the 20, 15 showed a positive bias (the roadside count was the larger of the two).*

The average difference for the pooled data of the latter 15 sets was +3.3 passengers per bus. As one would expect, when the data in the remaining 69 sets are pooled and tested, the result is not significant. The average difference for those counts was -0.4. The 20 significant tests are about twice as many as one would expect by chance alone with a 0.1 level of significance. The net implication of the evidence is that, while the overall results are unbiased, the counting procedures were "out of control," in the statistical sense, some part of the time. As we noted earlier, this finding led to the design of better controls for later surveys.

*One cannot say for certain whether this was caused by roadside overcounting or onboard undercounting. To the extent that our assumption is correct that the conditions in which onboard counting is done make it inherently more reliable, then the cause was roadside overcounting.

TABLE 1
SUMMARY STATISTICS

Differences (Roadside Minus Onboard) of Matched Pairs of Counts

Date Station Transit Line(s)	Time# Period	Sum of Differ- ences	Sum of Squares	Number of Matched Pairs	Average Differ- ence	Standard Deviation	Computed t-value
4/11/73	I-AM	13	37	6	2.2	1.33	4.0*
39	I-Off	4	26	7	0.6	1.99	0.8
71,72	I-PM	-8	32	9	-0.9	1.76	-1.5
	O-AM	25	109	8	3.1	2.10	4.2*
	O-Off	9	95	6	1.5	4.04	0.9
	O-PM	10	214	9	1.1	5.04	0.7
4/11/73	I-AM	-24	520	6	-4.0	9.21	-1.1
40	I-Off	16	120	7	2.3	3.73	1.6
N	I-PM	7	35	3	2.3	3.06	1.3
	O-AM	-10	138	3	-3.3	7.23	-0.8
	O-Off	20	202	6	3.3	5.20	1.6
	O-PM	-22	268	5	-4.4	6.54	-1.5
4/11/73	I-AM	-20	204	6	-3.3	5.24	-1.6
41	I-Off	-9	47	8	-1.1	2.30	-1.4
6,66	I-PM	3	55	9	0.3	2.60	0.4
	O-AM	7	177	8	0.9	4.94	0.5
	O-Off	-2	234	8	-0.3	5.78	-0.1
	O-PM	13	69	8	1.6	2.62	1.8
4/12/73	I-AM	7	29	2	3.5	2.12	2.3
42	I-Off	7	59	6	1.2	3.19	0.9
37	I-PM	4	22	5	0.8	2.17	0.8
	O-AM	-	-	1	-	-	-
	O-Off	13	175	8	1.6	4.69	1.0
	O-PM	6	42	4	1.5	3.32	0.9
4/12/73	I-AM	17	229	2	8.5	9.19	1.3
43	I-Off	10	42	3	3.3	2.08	2.8
73,44	I-PM	5	15	4	1.3	1.71	1.5
	O-AM	3	9	2	1.5	2.12	1.0
	O-Off	7	17	3	2.3	0.58	7.0*
	O-PM	4	78	3	1.3	6.03	0.4

I=Inbound, O=Outbound, AM=AM Peak, Off=Offpeak, PM=PM Peak.

* Significant at 0.1 level, 2-tail test

TABLE 1 (Continued)

Date Station Transit Line(s)	Time# Period	Sum of Differences	Sum of Squares	Number of Matched Pairs	Average Difference	Standard Deviation	Computed t-value
4/12/73	I-AM	-1	1	2	-0.5	0.71	-1.0
45	I-Off	8	20	6	1.3	1.37	2.4*
37,44	I-PM	1	31	5	0.2	2.77	0.2
	O-AM	4	8	3	1.3	1.15	2.0
	O-Off	-9	35	6	-1.5	2.07	-1.8
	O-PM	-7	39	6	-1.2	2.48	-1.2
4/24/73	I-AM	-13	89	7	-1.9	3.29	-1.5
44	I-Off	33	259	5	6.6	3.21	4.6*
K,L,M	I-PM	-27	281	6	-4.5	5.65	-2.0
	O-AM	-25	253	6	-4.2	5.46	-1.9
	O-Off	2	214	6	0.3	6.53	0.1
	O-PM	-20	234	4	-5.0	6.68	-1.5
4/24/73	I-AM	-17	743	11	-1.5	8.47	-0.6
46	I-Off	22	286	9	2.4	5.39	1.4
10,26	I-PM	3	257	9	0.3	5.66	0.2
	O-AM	-11	189	7	-1.6	5.35	-0.8
	O-Off	-15	243	12	-1.3	4.52	-1.0
	O-PM	2	182	9	0.2	4.76	0.1
4/25/73	I-AM	15	211	11	1.4	4.37	1.0
47	I-Off	6	148	23	0.3	2.58	0.5
12,14	I-PM	-4	238	18	-0.2	3.73	-0.3
	O-AM	35	227	8	4.4	3.25	3.8*
	O-Off	68	1032	24	2.8	6.04	2.3*
	O-PM	-7	105	15	-0.5	2.70	-0.7
4/25/73	I-AM	-25	825	8	-3.1	10.33	-0.9
50	I-Off	-38	372	4	-9.5	1.91	-9.9*
17	I-PM	-	-	0	-	-	-
	O-AM	4	66	6	0.7	3.56	0.5
	O-Off	5	13	2	2.5	0.71	5.0*
	O-PM	-	-	0	-	-	-
4/26/73	I-AM	3	317	9	0.3	6.28	0.2
54	I-Off	67	593	20	3.4	4.40	3.4*
10	I-PM	8	86	10	0.8	2.97	0.9
	O-AM	21	93	10	2.1	2.33	2.8*
	O-Off	25	147	15	1.7	2.74	2.4*
	O-PM	-19	159	7	-2.7	4.23	-1.7

TABLE 1 (Continued)

Date Station Transit Line(s)	Time# Period	Sum of Differences	Sum of Squares	Number of Matched Pairs	Average Difference	Standard Deviation	Computed t-value
4/26/73	I-AM	-10	158	5	-2.0	5.87	-0.8
56	I-Off	-2	52	5	-0.4	3.58	-0.3
25	I-PM	-6	338	5	-1.2	9.09	-0.3
	O-AM	7	17	5	1.4	1.34	2.3*
	O-Off	1	175	6	0.2	5.91	0.1
	O-PM	29	493	5	5.8	9.01	1.4
4/26/73	I-AM	0	132	6	0.0	5.14	0.0
53	I-Off	-1	1	8	-0.1	0.35	-1.0
30X	No other observations						
5/1/73	I-AM	39	341	6	6.5	4.18	3.8*
58	I-Off	12	144	2	6.0	8.49	1.0
15,42	I-PM	-	-	0	-	-	-
	O-AM	-20	126	5	-4.0	3.39	-2.6*
	O-Off	6	36	5	1.2	2.68	1.0
	O-PM	-	-	0	-	-	-
5/1/73	No Inbound Observations						
61	O-AM	29	235	15	1.9	3.56	2.1*
30X	O-Off	-2	28	10	-0.2	1.75	-0.4
	O-PM	-	-	0	-	-	-
5/1/73	O-Off	-5	25	6	-0.8	2.04	-1.0
62	No Other Observations						
17X							
5/3/73	I-AM	157	1957	29	5.4	6.29	4.6*
47	I-Off	0	498	16	0.0	5.76	0.0
12,14	I-PM	-34	368	22	-1.5	3.88	-1.9*
	O-AM	-40	860	24	-1.7	5.87	-1.4*
	O-Off	-4	212	14	-0.3	4.03	-0.3
	O-PM	-23	273	20	-1.2	3.60	-1.4*

2. Duplicate Onboard Counts

Duplicate onboard counts were made by pairs of observers at three different stations. Data from two of the stations were analyzed. Ridership at the third station was so sparse that it was not a good test of the methodology, and visual inspection of the data was sufficient to confirm a negligible difference between counters.

Counts were made at Station 44, on the K, L and M streetcar lines, and at Station 47, on the numbers 12 and 14 trolley bus lines. The summary statistics are given in Table 2.

Almost without exception, when there were standees, the observers saw all the seats occupied and recorded the same number of seated persons. Consequently, separate analyses were made of counts of seated passengers when there were no standees, and counts of standees when all seats were occupied.

For seated passengers, the mean differences were -0.1 in one set of data and 0.2 in the other, neither significantly different from zero. The corresponding standard deviations were 2.62 and 2.06.

Counts of standees, when a small number of outliers are removed, result in mean difference of 0.6 and 0.2,

TABLE 2

SUMMARY STATISTICS

Differences of Duplicate Onboard Counts

Date Station Transit Line(s)	Type of Count	Sum of Differences	Sum of Squares	Number of Matched Pairs	Average Difference	Standard Deviation	Computed t-value
4/24/73	(1)	-1	69	11	-0.1	2.62	-0.1
44	(2)	-25	455	9	-2.8	6.95	-1.2
K,L,M	(3)	4	22	7	0.6	1.81	0.8
	(5)	3	91	18	0.2	2.31	0.3
4/25-5/3/73	(1)	6	120	29	0.2	2.06	0.5
47	(2)	-13	201	7	-1.9	5.41	-0.9
12,14	(4)	1	5	6	0.2	0.98	0.4
	(5)	2	104	33	0.1	1.79	0.2

(1) Seated passengers, one or more empty seats

(2) Standing passengers, all data

(3) Standing passengers, less two outliers

(4) Standing passengers, less one outlier

(5) Combined seated and standing passengers, less outliers

respectively; and in standard deviations of 1.81 and 0.98. These parameters are comparable to those observed for seated passengers.

When the differences of total (seated and standing) counts are analyzed, not including extreme data, the averages at the two stations are 0.2 and 0.1, and the standard deviations are 2.31 and 1.79. None of the t-values in Table 2 is significant.

3. Measures of Counting Errors

After the evaluation of the counting methods, the second objective was to obtain estimates of counting errors and data variability so that required sample sizes could be determined for detecting changes in patronage of various magnitudes. To arrive at the estimates, we note the following:

(1) The estimates of the standard deviation of the differences of paired onboard counts from the two sets of data in Table 2 were 2.31 and 1.79. Taking the average of the squares, 4.26, gives twice the variance of the onboard counting procedure. Thus, our best estimate of onboard counting variance is 2.13 (s.d. = 1.46).

(2) The weighted average of the variances of the differences of all paired roadside/onboard counts was 20.67 (s.d. = 4.55). This variance is the sum of the variances of the two counting procedures.

Thus, we can obtain an estimate of the variance of the roadside procedure alone by the difference of (1) and (2), namely 20.67 - 2.13 = 18.54* (s.d. = 4.31). There is also an estimate of the variance of the AM peak count (Table 3, Appendix A), 19.4^2 = 376.36, which is itself the sum of the variances of the roadside counting procedure and of between-bus ridership. Then an estimate of the variance of AM peak between-bus ridership, again by difference, is 376.36 - 18.54 = 357.82 (s.d. = 18.9).

Now we see the relative magnitudes of the standard deviations of the onboard counting process (1.46), the roadside counting process (4.30), and between-bus variation (18.9). What is of particular interest to us is that the roadside counting process increases the observed between-bus standard deviation (through the root-mean-square) from 18.9 to 19.4, certainly

*It would have been preferable to obtain a direct measure of this quantity by doing some replication of roadside counts, as was done with the onboard counts. Arrangements were made for roadside count replication in the subsequent sampling period, but those data have not been analyzed yet.

a modest increase. Looking at it another way, if only onboard counts had been made, the observed between-bus standard deviation would have been reduced from 19.4 to 19.0.

We conclude that, as regards the standard deviation of the overall counting process, and therefore the ability to detect small changes in ridership, there is no practical difference between the two counting methods.

4. Sample Sizes and Detectable Differences

In Appendix A, the method for determining sample sizes required to detect changes of various magnitudes is described in detail. In summary, it was found that, based on the observed variability of the data in the San Francisco survey, for total ridership in a peak period, on a single street or highway crossing the screen line:

(1) To detect a change between two observation periods as small as 10 per cent, five-day samples taken in each period would be required.

(2) Two-day samples could, on the average, detect changes no smaller than 16 per cent.

IV. IMPACT ANALYSIS AND FUTURE DATA COLLECTION

A. Sample Sizes

There are several considerations besides statistical ones in deciding how large future samples should be, and they are of course interrelated. One of these is the matter of allocating a travel behavior study budget among the various kinds of measurement programs. Given the budgets that have been available so far, samples of at most two days at a station have been possible. Is there any way to reconcile this limitation with the evident need for samples of five and more days to detect relatively small changes? Yes, there is, if samples at several stations can be meaningfully pooled. In that case, two-day samples at each of three stations, say, would be the equivalent of six days of sampling.

Stations selected for pooling would have to be chosen carefully. A bus route may serve as a feeder line or as a line haul (between outlying district and CBD). Some lines serve as both, between different points on the route, and to different final destinations. The route networks of the major bus systems in the BART area are undergoing phased changes, coordinating with BART service. The changes involve increased or reduced service frequencies, and route changes. Thus, a bus line that is resurveyed at a given station after full BART service begins may be serving a different

function than it did before. With these cautions in mind, it is recommended that, as a first approximation, lines going through a counting station should be identified as feeders or line hauls and be pooled and analyzed accordingly.

B. Quality Control

Analysis of the San Francisco transit count data has convinced us that when "good" data are obtained, the roadside counting methodology does not increase observed between-bus variability from actual variability to an unacceptable extent. The questions of what are good data and how to get them are the crux of the matter, and the answers lie in quality control--control of field procedures, control of data management, and feedback between the two. It is not sufficient to "calibrate" a transit count procedure in a given time and place, and to rely on that calibration forever. Count surveys are far enough apart in time that each time one is done, new crews have to be recruited and trained. Quality control procedures, beyond ordinary field supervision, have to be built in and applied continually. Such procedures were implemented for the survey subsequent to that reported on here, and will be recommended, with appropriate modifications, for future surveys. The augmented controls included:

1. Design of field procedures. Onboard counting checks were made again, and in some instances duplicate roadside

counts were also made. The survey took place in an area served by AC Transit, which occasionally does check counts of its own with personnel who are more experienced than our field crews. AC Transit assigned one of their counters to do check counts at times and locations selected by MTC, to serve as an additional control on our procedures. Whereas in San Francisco each location (with one exception) was surveyed for only one day, the recent survey counted at each location on two days, at least a week apart.

2. Data Management procedures. Raw data were examined at the end of each day by a crew of experienced checkers. Missing data and errors were identified and, if recoverable, corrected before key-punching. Corrections were made from the context, if possible, or by checking with the field crews. Cards were punched within a week of receipt of the data, and listings were examined by the checkers, at which time additional errors were found. Errors that arose from misunderstanding or misapplication of proper field procedures were identified, and the need for corrective action was communicated to field supervisors. Whenever possible, this was done in time for the following day's operations, and in cases where data were ambiguous, field crews were queried while their memories were still fresh.

A third type of control, routinized and short-delay analysis of control counts, has yet to be implemented.

With the planned programs of active quality control and analysis of pooled data, it is expected that the transit passenger counts will satisfy the analytical requirements of the BART Impact Program.

APPENDIX A

A Model for Detection and Estimation of Changes in Transit Patronage

APPENDIX A

A Model for Detection and Estimation of Changes in Transit Patronage

The formulation of an analytical model relies on fundamental theory, in this case statistical, but at the same time it demands a practical outlook. The model has to reflect things that we know or suspect in a qualitative way, such as:

(1) Transit ridership peaks twice a day, but the shape of the peaks is not necessarily a flat plateau, and their beginning and end will vary from day to day.

(2) Even in a period of relatively level demand, there may be considerable variability in occupancy between vehicles; every transit rider is familiar with the phenomenon of the jammed, behind-schedule bus followed by a relatively uncrowded on-time vehicle.

(3) Many factors will cause deviations from schedules, so that any counting method that observes ridership on a bus route between two specific clock times on different days, may be observing different numbers of buses, and hence different fractions of ridership.

(4) The time distribution of ridership is different on mid-week days than it is on Mondays and Fridays, and field sampling should take this into account.

(5) Within a group, say of midweek days, there is day-to-day variability. Sampling plans also have to take into account seasonal variations.

(6) A single day's data may reflect extraordinary circumstances--accidents, bad weather, unusual events.

(7) In aggregating data from more than one route, the analyst has to be alert to the possibility that the different routes are affected differently by the innovation or change whose impact is being studied, and hence it may be misleading to aggregate those data.

For example, we may be interested in observing whether there has been a change in patronage on Muni lines paralleling BART, from sometime before to sometime after the start of BART service, and estimating what the magnitude of that change might be. If we choose our measures carefully, the analysis itself can be simple, namely a t-test and the computation of confidence limits using the t-distribution.

An obvious measure is total ridership per day observed crossing the screen line at a given station, say inbound from 6-9 AM--the AM peak.

An alternative measure is the average number of riders per bus, in the same period. If the assumption is justified that the individual counts are random observations from a normal distribution, then the average per bus is preferable to the average daily total because it has better analytical properties. This

TABLE 3

SUMMARY STATISTICS

Roadside Counts, Duplicate Days

	Date		
	4/25	5/3	Total/Pooled
Individual bus counts			
Number	61	65	126
Sum	2,218	2,547	4,765
Average	36.4	39.2	37.8
s.d.	19.7	19.2	19.4
skewness			0.25
kurtosis measure			2.48
Daily AM Peak Total			
Number of days			2
Sum	2,218	2,547	4,765
Average			2,383
s.d.			233
Daily AM Peak Total Normalized for Equal Count (63)			
Number of days			2
Count	2,291	2,469	4,759
Average			2,380
s.d.			126

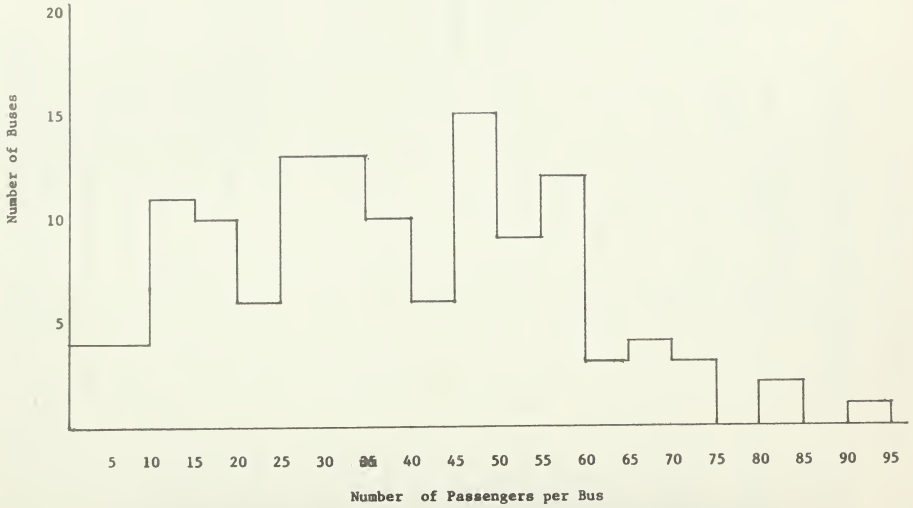
will become clear in the example to be given.

Curbside counts made on April 25 and May 3, 1973 during the AM peak periods, inbound, were analyzed. The location was station 47, and the vehicles were trolley buses on the Nos. 12 and 14 lines. The summary statistics are shown in Table 3. Looking at the individual bus counts first: although the average count per bus was almost 8 per cent higher on the second day than on the first, the standard deviation was remarkably stable between days, and the difference between daily averages was not significant. A histogram of the pooled data, Fig. 2, showed a worrisome amount of variation between classes, but tests of departure from normality were reassuring. The skewness coefficient, 0.25, was not significantly different from zero. The kurtosis measure was borderline, confirming the visual impression from the histogram of some deficit of probability in the tails of the distribution. Thus, proceeding with caution, one is willing to use the t-test, which requires underlying normality.

The need for caution is illustrated by the distribution of some of the Seattle data, the differences of 53 matched pairs of on-board-curbside counts during the PM peak period, outbound, all routes pooled. It was reported that the average difference was 0.1, and a t-test failed to find this significantly different from zero, leading to a conclusion of no bias in the curbside counting method. However, an examination of the data plots in the Seattle report led us to test the assumption of normality.

Fig. 2

Histogram of Transit Passenger Counts, Station 47, Two Days Combined



The skewness coefficient for those data was -1.98, indicating a long tail to the left, and hence a selective bias--relatively small positive differences for 26 observations, and a wider scatter of 14 negative differences (the remaining differences were zero). In this case, the t-test is inappropriate.

Returning to the San Francisco data, consider the daily totals for the AM peak period. Because different numbers of buses were observed, 61 on one day, 65 on the other, the standard deviation between days, 233,is greater than it should be. Digging back through block number (scheduled run number) data, one sees why, in terms of deviations by individual buses from schedule. For example, some buses that passed the screen line before 9 AM on one day did not arrive till after 9 on the other. This source of variation can, and should be, eliminated. We chose to do this by averaging the number of bus runs and weighting the data accordingly, multiplying the first day's count by 63/61 and the second day's by 63/65. The average remains almost the same, but the standard deviation between days is reduced to 126.

Now the relationships between sample sizes and the magnitudes of detectable differences can be explored. Since a large number of comparisons of ridership changes will be made eventually, we can accept a relatively high level of Type I error* in individual cases, namely 0.1**. Also, we should always be prepared

---

*Probability of rejecting the null hypothesis when it is true.

**The choice of this level creates some problems in determining Type II error. Whereas operating characteristic curves are readily available for smaller levels of error, values for 0.1 had to be computed. See Appendix B.

to test no change against the alternative that ridership has changed one way, so a one-tail test is appropriate.

Using as the test statistic the average count per bus, and the observed sample over two days of 126 buses, we consider three cases:

(1) If the second sample, against which a comparison will be made, is also 126 buses, and we test no change against the altnerative that there has been a 10 per cent decrease in ridership, what is the Type II error*, assuming the mean and s.d. (standard deviation) of Table 3? It is 0.38, which is rather high.

(2) Testing against the same alternative, a 10 per cent decrease, how large a sample is required in each sampling period to assure a Type II error of no more than 0.1? The answer is 320, which would require 5 days of sampling.

(3) Finally, for the 2-day sample and a Type II error of 0.1, how small a change can be detected? The answer is 16 per cent.

Now considering the corresponding cases for the average daily AM peak period total:

1. Given two samples of two days each, and testing for no change, against the 10 per cent alternative, the Type II

---

*Probability of accepting the null hypothesis against a specific alternative when it is false.

error is 0.49.

2. Testing against the same alternative, and given samples of 5 days each, the Type II error is about 0.12.

3. Given samples of two days each, and testing against a 16 per cent alternative, the Type II error is 0.24.

Because of computational difficulties for small sample sizes (see Appendix B), the Type II error could not be specified as input for the last two cases, but rather was computed as a function of sample size and the magnitude of hypothesized change.

Comparing the Type II errors between the test of the means per bus and the mean daily totals, it is seen that for the cases where the differences of two-day samples are tested, the errors associated with tests of the means per bus are the smaller, by a substantial amount. For the larger samples of 5 days each, the bus-mean error is 0.10, whereas the daily-mean error is 0.12, not much of a difference.

The foregoing examples illustrate that, if individual bus counts can be justifiably treated as samples from a normal distribution, the measure to be tested should be the mean of those counts. However, if there are significant departures from normality, the average daily total can nevertheless be treated as a sample from a normal distribution (because of the Central Limit Theorem), and should be used, provided a large enough sample can be taken.

APPENDIX B

Methods for Computing Type II Errors

APPENDIX B

METHODS FOR COMPUTING TYPE II ERRORS

An OC (operating characteristic) curve gives the probability, β, of failing to reject the null hypothesis when a particular alternative is true and a random sample of a given size is used to test the hypothesis. The probability, β, is also called the type II error. OC curves for the one-sided t-test, and levels of significance 0.05 and smaller are readily available, for example in Bowker and Lieberman (1).

Values of β for larger levels of significance are more difficult to find. Methods for computing them are given in Ferris et al (2). Ferris gives two equations for computing β. The first is for large sample sizes (say greater than 10), and was used to compute the errors associated with changes in the mean count per bus. It is described as a good approximation to the exact value, and is

$$\beta = \Pr\{ -t_\alpha + d\sqrt{n} < t < t_\alpha + d\sqrt{n} \} \qquad (1)$$

This is a simple equation to evaluate. In the one-sided case, t_α is the tabulated value of t with n-1 degrees of freedom, and is the value such that a t-variate exceeds it with probability α (the level of significance). The sample size, m, for each of the samples whose means are being compared is given by

$$n = 2m - 1 \quad . \qquad (2)$$

The value d is the ratio $\Delta/2\sigma$, where Δ is the magnitude of the difference one wishes to detect, and σ is the standard deviation of the observed distribution. Substituting the values of t_α, d and n in the inequalities in braces in Eq.(1), the expression represents the probability that a t-variate with n-1 d.f. is within those limits. The expression is easily evaluated, using tables of the t-distribution.

For smaller samples, Ferris recommends the equation

$$\beta = \exp(-\tfrac{1}{2}nd^2) \sum_{r=0}^{\infty} \frac{(\frac{1}{2}nd^2)^r}{r!} \, I\left[(r+\tfrac{1}{2}), \tfrac{1}{2}(n-1); \frac{t_\alpha^2}{n-1+t_\alpha^2} \right] \qquad (3)$$

This equation was used to compute errors for the per-peak period totals.

---

(1) op. cit.

(2) Charles L. Ferris, Frank E. Grubbs, Chalmers L. Weaver, "Operating Characteristics for the Common Statistical Tests of Significance," Ann. Math. Stat., June, 1946.

The symbols are as previously defined, with the addition that r is an index and I is the Incomplete Beta Function, with the three expressions in brackets representing its parameters, a, b and x. Formidable as the expression may seem, it converges rapidly, and is readily evaluated with the aid of a simple computer program. Values of I do have to be input to the program, but they can be obtained easily with the aid of published curves (3).

(3) E. S. Pearson and O.H. Hartley, eds., Biometrika Tables for Statisticians, Cambridge University Press, 1954. Cf. Table 17, p. 156.

APPENDIX C

Written Instructions to Transit Count Teams

MEMO TO: Special Transit Count Team (Onboard Counters)

FROM: Joel Markowitz

SUBJECT: Instructions for Conducting Count

I would like to repeat the guidelines I gave you last week on how we would like this special study to be conducted.

At each location where regular transit count crews are located, according to the schedule you have been given, a morning and an afternoon crew from your team will be boarding Muni buses just before and leaving just after the buses (or streetcars or trolleys) cross the screen-line where the regular road-side transit crew is positioned. Your field survey of Tuesday will give you information on just where the best stops are for this purpose.

An exact head count, rather than an estimate, should be made. For example, counting standees only may be inaccurate if some of the seated passengers are occupying more that one seat with packages. Please make sure your assumption of "all seats taken" is true before you proceed to count standees only.

If there are no standees, please make a full head count. In other words, do not take for granted that the "capacity" figure we have on record for the type of vehicle is precise. Remember: you are being counted upon to provide the most accurate transit counts. If we can't be certain that you have taken all steps

to be 100% accurate, our analysis comparing your data to those collected by road-side teams will fail.

Your procedures should be:

1) Know exactly where the road-side team is located, and where your boarding and leaving stops are each day.

2) Know which Muni routes you are to board each day.

3) Have tokens ready when boarding.

4) Do not identify yourself to operator or passengers unless asked; then be honest, but brief so as not to interfere with your task.

5) Do not be conspicuous while counting passengers. Count while moving through the vehicle after boarding on your way to your best vantage point. Do not count out loud or by pointing your finger, pen, or pencil at passengers.

6) Make sure you have an accurate count at the moment the vehicle passes the road-side crew.

7) Do not be obvious in recording your data. If possible, record data on a small note pad, then transfer to the large standard forms after leaving the vehicle and before boarding the next.

8) If any peculiarities are encountered in taking Muni (any more than usual?), please note them in a daily diary and give to the team captain. For instance, if the bus you catch doesn't stop where you thought it would, or takes a different route and doesn't cross the screen-line, bring these facts to the attention of the team captain as soon as possible.

9) Do not wear the MTC identification badge; carry it in pocket or purse.

10) You will be issued one roll of Muni tokens per day. (20 tokens) Use them as follows:

6 AM - 7 AM -- take the first vehicle after the hour and after the half-hour; i.e., two observations

7 AM - 9 AM -- take every vehicle possible each direction; i.e., take the first vehicle that arrives, count passengers, get off, take the next vehicle the other direction, etc. Probably no more than four or six per hour.

9 AM - end of shift -- same as 6 AM - 7 AM

noon - 4 PM -- same as 6 AM - 7 AM

4 PM - 6 PM -- same as 7 AM - 9 AM

6 PM - end of shift -- same as 6 AM - 7 AM

11) If you have extra tokens, return them to the team captain; if you run short, report this to the team captain as soon as possible. This first week we may be under or over-estimating how much is needed.

12) Please pick up a transfer each time you board a vehicle. These, with your record sheets turned in to the team captain, will verify how many vehicles you were able to board each day.

13) Work out with the team captain how breaks will be scheduled.

14) Do not compare notes and data with road-side crew members. To make your data useful for analysis, we have to be certain that your observations are totally independent from those of the road-side crews.

15) On Thursday, April 12 deviate from the normal procedure in this way for that day only: board all vehicles that day in pairs, so that the independent observations of two of you for that vehicle can be compared. We mentioned before that under peak load conditions, an onboard observer might not be able to make an accurate count; we want to know how accurate. So on this day, two of you will count all people on each of the buses: do not compare figures. Work out with your team captain which two of the three persons on the crew will form this pair.

If you have any other questions, please call me at MTC, 849-3223, any time between 8:30 and 5:00.

Good luck.

TRANSIT OCCUPANCY COUNT PROCEDURE

(Roadside Counters)

Place yourself so that you are able to see all information on front of bus (i.e., vehicle number, destination, etc.) and the name of operating agency of each transit vehicle.

Observe and record time, route number, vehicle number, company code, block number, and destination for the transit vehicles. In addition, observe and record occupancy of regularly operated mass transit which stop at your count location or sufficiently slow moving to allow occupancy estimation.

All information can be recorded on the forms provided; block number will precede the destination description. Write total number of passengers seated under the "seated" column; then put down standees under "standing" if there are any passengers standing.

Occupancy counting will be made by one of three ways as indicated below:

1. If a vehicle is less than one-half full, count number of passengers, OR

   If you have an actual count of the passengers, even if there are standees, enter that number in the passengers seated column.

2. If a vehicle is more than one-half full, count number of vacant seats. If the vehicle is full with no standees, enter 000 in the vacant seats column.

3. If there are standees, count number of standees. It is assumed that all seats are occupied for this instance.

Observation should be made at least from 6:00 AM to 7:00 PM. Counting should be continuous during the count day. Do not fill in unnecessary leading zeros. If you need help, ask your captain for help.